

MEMORY CONTROLLER AND CONTROL METHOD

Patent Number: JP2001243206

Publication date: 2001-09-07

Inventor(s): HANEDA MITSUMASA; NISHIYAMA SHINJI; OGAWA YUICHI; HANAOKA YUJI; TOMIHARI KIYOTAKA

Applicant(s): FUJITSU LTD

Requested Patent: ☐ JP2001243206

Application Number: JP20000050925 20000228

Priority Number (s):

IPC Classification: G06F15/177; G06F12/16; G06F13/00; G06F13/16; H04L13/08

EC Classification:

Equivalents:

Abstract

PROBLEM TO BE SOLVED: To notify a reply packet received by completion of packet transfer to an access origin by bus transfer by performing the packet transfer of a write command after performing the bus transfer of it.

SOLUTION: A verification command generation control part 36 is provided in a processing module 20 on the host side and a verification command replying part 38 is provided in a bridge module 24. The verification command generation control part 36 issues a verification command to verify a state of the packet transfer of the write command by a packet network after normally completing transfer of an interface bus 12 by issuing a write command to a memory module 32. The verification command replying part 38 replies to a replay packet being received corresponding to the packet transfer of the write command as read data and makes the replay packet recognize whether or not data writing is normally performed when the verification command is received.

Data supplied from the esp@cenet database - I2

a)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-243206

(P2001-243206A)

(43) 公開日 平成13年9月7日 (2001.9.7)

| (51) Int.Cl. ⁷ | 識別記号 | F I | テ-アコ-ト*(参考) |
|-------------------------------|-------|----------------|-------------------|
| G 0 6 F 15/177 | 6 7 6 | G 0 6 F 15/177 | 6 7 6 A 5 B 0 1 8 |
| 12/16 | 3 3 0 | 12/16 | 3 3 0 D 5 B 0 4 5 |
| 13/00 | 3 0 1 | 13/00 | 3 0 1 J 5 B 0 6 0 |
| 13/16 | 5 2 0 | 13/16 | 5 2 0 B 5 B 0 8 3 |
| H 0 4 L 13/08 | | H 0 4 L 13/08 | 5 K 0 3 4 |
| 審査請求 未請求 請求項の数 8 O L (全 19 頁) | | | |

(21) 出願番号 特願2000-50925(P2000-50925)

(22) 出願日 平成12年2月28日 (2000.2.28)

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番1号

(72) 発明者 羽根田 光正

神奈川県横浜市港北区新横浜2丁目4番19号 株式会社富士通プログラム技研内

(72) 発明者 西山 真治

神奈川県横浜市港北区新横浜2丁目4番19号 株式会社富士通プログラム技研内

(74) 代理人 100079359

弁理士 竹内 進 (外1名)

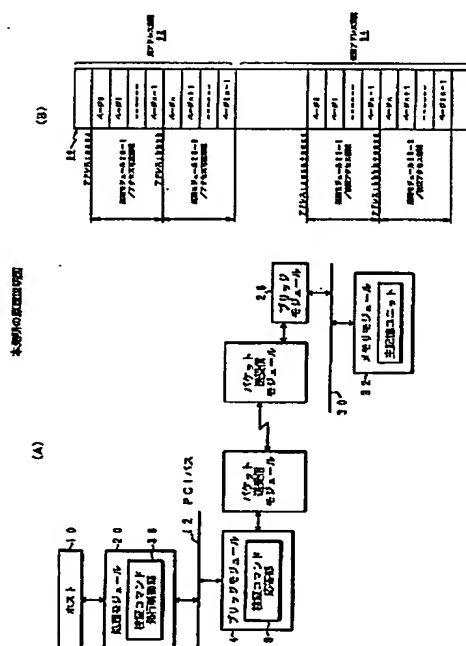
最終頁に続く

(54) 【発明の名称】 メモリ制御装置及び制御方法

(57) 【要約】

【課題】 ライトコマンドをバス転送した後にバケット転送し、バケット転送完了で受信したリブライバケットをバス転送にによってアクセス元に通知できるようにする。

【解決手段】 ホスト側の処理モジュール20に検証コマンド生成制御部36を設け、ブリッジモジュール24に検証コマンド応答部38を設ける。検証コマンド生成制御部36は、メモリモジュール32にライトコマンドを発行してインタフェースバス12の転送を正常終了した後に、バケットネットワークによるライトコマンドのバケット転送の状態を検証する検証コマンドを発行する。検証コマンド応答部38は、ブリッジモジュール24に、検証コマンドを受信した際に、ライトコマンドのバケット転送に対応して受信しているリブライバケットをリードデータとして応答してデータ書込が正常に行われたか否かを認識させる。



【特許請求の範囲】

【請求項1】第1のモジュールを接続したインタフェースバスと第2のモジュールを接続したインタフェースバスを、各々ブリッジモジュールを介してバケットネットワークで接続し、前記第1のモジュールから第2のモジュールにライトコマンド又はリードコマンドを発行して書き込み又は読み出しを実行するネットワーク構成のメモリ制御装置に於いて、

前記第1のモジュールに、前記第2のモジュールにライトコマンドを発行してインタフェースバスの転送を正常終了した後、前記バケットネットワークによるライトコマンドのバケット転送状態を検証する検証コマンドを発行する検証コマンド生成制御部を設け、

前記ブリッジモジュールに、前記検証コマンドを受信した際に、前記ライトコマンドのバケット転送に対応して受信しているリブライバケットをリードデータとして応答してデータ書き込みが正常に行われたか否かを認識させる検証コマンド応答部を設けたことを特徴とするネットワーク構成のメモリ制御装置。

【請求項2】請求項1記載のネットワーク構成のメモリ制御装置に於いて、前記検証コマンド生成部は、前記ライトコマンドのアドレスに所定のオフセットを加算した実アドレス空間を越えるアドレスを指定したリードコマンドを検証コマンドとして発行することを特徴とするネットワーク構成のメモリ制御装置。

【請求項3】請求項1記載のネットワーク構成のメモリ制御装置に於いて、前記ブリッジモジュールは、更に、前記ライトコマンドによるバケットの転送中にあることを登録する転送中登録部を有し、前記コマンド応答部は、前記検証コマンドを受信した際に前記転送中登録部を参照して転送中であればリトライコマンドを応答し、転送終了であればリブライバケットをリードデータとして応答することを特徴とするネットワーク構成のメモリ制御装置。

【請求項4】請求項1記載のネットワーク構成のメモリ制御装置に於いて、前記インタフェースバスはPCI (Peripheral Component Interconnect)バスであり、前記ブリッジモジュールはPCIブリッジモジュールであることを特徴とするネットワーク構成のメモリ制御装置。

【請求項5】第1のモジュールを接続したインタフェースバスと、第2のモジュールを接続したインタフェースバスを各々ブリッジモジュールを介してバケットネットワークで相互に接続し、前記第1のモジュールから第2のモジュールにライトコマンド又はリードコマンドを発行して書き込み又は読み出しを実行するメモリ制御方法に於いて、

前記第1のモジュールから前記第2のモジュールにライトコマンドを発行してインタフェースバスの転送を正常終了した後、前記バケットネットワークによるライトコマンドのバケット転送の状態を検証する検証コマンド

を発行し、

前記ブリッジモジュールで、前記検証コマンドを受信した際に、前記ライトコマンドのバケット転送に対応して受信しているリブライバケットをリードデータとして応答してデータ書き込みが正常に行われたか否かを認識させることを特徴とするメモリ制御方法。

【請求項6】請求項5記載のネットワーク構成のメモリ制御装置に於いて、前記第1のモジュールは、前記ライトコマンドのアドレスに所定のオフセットを加算した実アドレス空間を越えるアドレスを指定したリードコマンドを検証コマンドとして発行することを特徴とするネットワーク構成のメモリ制御装置。

【請求項7】請求項5記載のメモリ制御方法に於いて、前記ブリッジモジュールは、前記ライトコマンドによるバケット転送中の有無を登録し、前記検証コマンドを受信した際に転送中であればリトライコマンドを応答し、転送終了を待ってリブライバケットをリードデータとして応答することを特徴とするメモリ制御方法。

【請求項8】請求項5記載のメモリ制御方法に於いて、前記インタフェースバスはPCI (Peripheral Component Interconnect)バスであり、前記ブリッジモジュールはPCIブリッジモジュールであることを特徴とするメモリ制御方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ホストや入出力デバイスに対応して設けた処理モジュールを、バケットネットワークを介して分散配置された主記憶として機能するメモリユニットに接続してメモリアクセスを実行するネットワーク構成のメモリ制御装置及び制御方法に関する、特に、バケット転送の状態をPCIバスで接続されたアクセス元のモジュールに通知できるようにしたメモリ制御装置及び制御方法に関する。

【0002】

【従来の技術】従来、複数のホストに対し主記憶として機能するメモリユニットを分散配置したメモリ制御システムにあっては、ホスト、メモリユニット及び入出力デバイスを高速拡張バスとして知られたPCIバス(Peripheral Component Interconnect Bus)で接続している。

【0003】一方、近年におけるCPUの処理性能の向上と高速化に伴ない、PCIバスに接続しているホスト、メモリユニット及び入出力デバイスをバケットネットワーク(バケットバス)で接続したシステムが提案されている。このようなバケットネットワークを使用して主記憶を分散したシステムとしては、例えばクラスタシステム等が知られている。

【0004】図1は、バケットネットワークでホストとメモリユニットを接続した主記憶制御装置である。ホスト000はモジュール202によりPCIバス204に接続され、PCIバス204はPCIブリッジモジュール

208
3 14 2 (3)

ール206によりパケットネットワーク210のパケット送受信モジュールに接続される。また主記憶ユニット220を備えたメモリモジュール218は別のPCIバス216に接続され、このPCIバス216もPCIブリッジモジュール206を介してネットワーク210のパケット送受信モジュール212に接続されている。

【0005】ホスト200からのライト命令があると、モジュール202がライトコマンド（コマンドコード、アドレス、データを含む）をブリッジモジュール206にバス転送する。このバス転送が完了すると、PCIブリッジモジュール206は、ライトコマンドをパケットデータに変換し、パケット送受信モジュール208からネットワーク210に転送し、パケット送受信モジュール212を介してPCIブリッジモジュール214で受信される。

3
【0006】PCIブリッジモジュール214はライトアクセスのパケットの受信を終了すると、正常に受信したか否かを示すリブライパケットを送信元のPCIブリッジモジュール206に返す。正常終了であればPCIブリッジモジュール214は、受信したライトアクセスのパケットから得たライトコマンドをPCIバス216によりメモリモジュール218に転送し、主記憶ユニット220に対するデータ書き込みが行われる。

【0007】

【発明が解決しようとする課題】しかしながら、このようなホストからパケットネットワークを介して主記憶にアクセスする装置にあっては、送信側のPCIブリッジモジュール206が送信先のPCIブリッジ214からのリブライパケットによって異常終了を受信した場合、転送元であるモジュール202に通知することができない問題があった。

【0008】即ち、PCIブリッジモジュール206は、モジュール202からのPCIバス204によるライトコマンドのバス転送が正常終了した後に、ネットワーク210に対するパケット転送を開始しており、PCIブリッジモジュール214から異常終了を示すリブライパケットを受信した際には、送信元のモジュール202とのバス接続が既に完了しており、リブライパケットによるパケット転送の異常終了を通知することができない。

【0009】本発明は、ライトコマンドをバス転送した後にパケット転送し、パケット転送完了で受信したリブライパケットをバス転送によってアクセス元に通知できるようにしたネットワーク構成の主記憶制御装置及び制御方法を提供することを目的とする。

【0010】

【課題を解決するための手段】図1は本発明の原理説明図である。

【0011】まず本発明は、図1(A)のように、ホスト10に対応して設けた処理モジュール（第1のモジュール）20を接続したインタフェースバス12とメモリモジュール（第2のモジュール）32を接続したインタフェースバス30を、各々ブリッジモジュール20、26を介してパケットネットワーク18で接続し、処理モジュール20からメモリモジュール32にライトコマンド又はリードコマンドを発行して書き込み又は読み出しを実行するネットワーク構成のメモリ制御装置を対象とする。

4

【0012】このようなネットワーク構成のメモリ制御装置として本発明は、処理モジュール20に検証コマンド生成制御部36を設け、ブリッジモジュール24に検証コマンド応答部38を設けたことを特徴とする。

【0013】検証コマンド生成制御部36は、メモリモジュール32にライトコマンドを発行してインタフェースバス12の転送を正常終了した後に、パケットネットワークによるライトコマンドのパケット転送の状態を検証する検証コマンドを発行する。検証コマンド応答部38は、ブリッジモジュール24で検証コマンドを受信した際に、ライトコマンドのパケット転送に対応して受信しているリブライパケットをリードデータとして応答してデータ書き込みが正常に行われたか否かを認識させる。

【0014】このためブリッジモジュール24でリブライパケットを受信した際に送信元の処理モジュール20とのバス転送が完了していても、送信元の処理モジュール20からライトコマンドに続いて発行された検証コマンドに対する応答としてリブライパケットの内容を送信元の処理モジュール20にインタフェースバス12を使用して通知することができ、もしリブライパケットが異常終了を示していれば、送信元の処理モジュール20からライトコマンドを再送する等の異常終了に対する適切な対応をとることができ、パケットネットワークを経由した主記憶アクセスの信頼性を向上することができる。

【0015】ここで処理モジュール20の検証コマンド生成制御部36は、ライトコマンドのアドレスに所定のオフセットを加算した実アドレス空間を越えるアドレスを指定したリードコマンドを検証コマンドとして発行する。

【0016】またブリッジモジュール24は、更に、ライトコマンドによるライトパケットの転送中にあることを登録する転送中登録部を有し、コマンド応答部38は、検証コマンドを受信した際に転送中登録部を参照して転送中であればリブライコマンドを応答し、転送終了であればリブライパケットをリードデータとして応答する。

h
【0017】インタフェースバス12は例えばPCI (Peripheral Component Interconnect)バス12であり、ブリッジモジュール24はPCIブリッジモジュール24である。

【0018】また本発明は、ネットワーク構成のメモリ制御方法を提供する。このメモリ制御方法は、処理モジュール20を接続したインタフェースバス12とメモリモジュール（第2のモジュール）32を接続したインタフェースバス30を、各々ブリッジモジュール20、26を介してパケットネットワーク18で接続し、処理モジュール20からメモリモジュール32にライトコマンド又はリードコマンドを発行して書き込み又は読み出しを実行するネットワーク構成のメモリ制御装置を対象とする。

ジュール20を接続したインタフェースバス12とメモリモジュール32を接続したインタフェースバス12を、各々ブリッジモジュール24を介してバケットネットワークで相互に接続し、処理モジュール20からメモリモジュール32にライトコマンド又はリードコマンドを発行して書き込み又は読み出しを実行するメモリ制御方法について、処理モジュール20からメモリモジュール32にライトコマンドを発行してインタフェースバス12の転送を正常終了した後に、バケットネットワークによるライトコマンドのバケット転送の状態を検証する検証コマンドを発行し；ブリッジモジュール24で、検証コマンドを受信した際に、ライトコマンドのバケット転送に対応して受信しているリブライバケットをリードデータとして応答してデータ書込が正常に行われたか否かを認識させる。

【0019】このネットワーク構成のメモリ制御方法の詳細は装置構成の場合と同じになる。

【0020】

【発明の実施の形態】図2は、本発明のメモリ制御装置及び制御方法が適用される計算機システムのシステム構成図である。

【0021】図2において、本発明が対象とする計算機システムはホスト10-1、10-2を有し、それぞれPCIバス12-1、12-2に接続されている。ホスト10-1を接続したPCIバス12-1には入出力サブシステムとして機能するデバイス14-1が接続され、ホスト10-2を接続したPCIバス12-2にも同様にデバイス14-2が接続される。

【0022】ホスト10-1、10-2の主記憶として機能するメモリユニットは、筐体16-1、16-2、・・・16-nのそれぞれに収納されており、筐体16-1～16-nのそれぞれはPCIバス12-1、12-2に接続される。これら筐体16-1～16-nはネットワーク（バケット転送バス）18で接続されており、ネットワーク18を経由して筐体16-1～16-nの相互間でバケット転送ができるようにしている。

【0023】図3は、図2の筐体16-1の内部構成を外部のホスト及びデバイス側と共に示している。尚、図2にあっては、PCIバス12-1、12-2は筐体16-1～16-nの外部バスとして示しているが、実際の装置にあっては図3のように、PCIバス12-1、12-2は筐体16-1内に位置しており、これに対し外部に設置しているホスト10-1、10-2はデバイス14-1、14-2をPCIバスのモジュールを介して接続することになる。

【0024】図3において、筐体16-1にはホスト10-1、10-2からのライト命令やリード命令に対応してライトコマンドやリードコマンドを発行する処理モジュール20-1、20-2が設けられている。処理モジュール20-1、20-2はPCIバス12-1、1

2-のそれぞれに接続される。またPCIバス12-1、12-2にはデバイスモジュール22-1、22-2を介して外部のデバイス14-1、14-2が接続されている。

【0025】一方、筐体16-1内には別のPCIバス30-1、30-2によりメモリモジュール32-1、32-2が接続されており、メモリモジュール32-1、32-2のそれぞれには主記憶ユニット34-1、34-2が設けられている。このような処理モジュール20-1、20-2を接続したPCIバス12-1、12-2と、メモリモジュール32-1、32-2を接続したPCIバス30-1、30-2は、バケットネットワークを介して接続される。

【0026】バケットネットワークは、処理モジュール20-1、20-2に対応して設けたPCIブリッジモジュール24-1、24-2、メモリモジュール32-1、32-2に対応して設けたPCIブリッジモジュール26-1、26-2、及びPCIブリッジモジュール24-1、24-2、26-1、26-2を接続したバケット送受信モジュール28で構成される。このバケット送受信モジュール28には更に、他の筐体例えば隣接する筐体16-2がネットワーク18で接続される。

【0027】このような筐体16-1の構成は、図2に示した他の筐体16-2～16-nについても同様である。ここでホスト10-1から処理モジュール20-1が例えばメモリモジュール32-2の主記憶ユニット34-2に対するライト命令を受けた場合の処理を簡単に説明すると次のようになる。

【0028】ホスト10-1からのライト命令に対し、処理モジュール20-1はPCIバス12-1によりPCIブリッジモジュール24-2に対しライトコマンドを発行する。このライトコマンドは、コマンドコード、アドレス、及び書込データを含む。

【0029】PCIブリッジモジュール24-1に対するライトコマンドのPCIバス12-1による転送が完了すると、PCIブリッジモジュール24-1はバケット送受信モジュール28を介して、アクセス先のメモリモジュール32-2に対応して設けているPCIブリッジモジュール26-2に対しライトコマンドのバケット転送を行う。

【0030】PCIブリッジモジュール26-2はライトコマンドのバケット転送が完了すると、送信元のPCIブリッジモジュール24-1に対しバケット転送の正常終了または異常終了を示すリブライバケットを送信する。

【0031】PCIブリッジモジュール26-2で正常にライトコマンドのバケット受信が完了した場合には、PCIバス32-2によりライトコマンドをメモリモジュール32-2に転送し、主記憶ユニット34-2の指定したアドレスに対するデータ書込みを行う。

【0032】この処理モジュール20-1からメモリモジュール32-2にライトコマンドを送ってデータ書込みを行う際のアドレス変換は、図4のようになる。図4は、図3の処理モジュール20-1から発行するライトコマンドのアドレス40-1、次のPCIブリッジモジュール24-1からパケット転送するライトコマンドのアドレス40-2、更にPCIブリッジモジュール26-2からメモリモジュール32-2に発行するライトコマンドのアドレス40-3を示している。

【0033】まず処理モジュール20-1からPCIブリッジモジュール24-1に出力されるライトコマンドのアドレス40-1は、上位ビット側からベースアドレス42、ノードID44及びMSUアドレスオフセット46で構成される。ベースアドレス42は、PCIバスに接続しているPCIブリッジモジュールを指定する。

【0034】ノードID44は、パケット送受信モジュール28に接続されている各PCIブリッジモジュール24-1、24-2、26-1、26-2にアサインされるID番号であり、この場合にはアクセス先のメモリモジュール32-2に対応して設けたPCIブリッジモジュール26-2のノードIDが設定される。更にMSUアドレスオフセット46は、メモリモジュール32-2に設けた主記憶ユニット34-2のメモリ空間のアドレスを示すオフセットである。

【0035】この処理モジュール20-1から出力されたアドレス40-1は、PCIブリッジモジュール24-1からパケット送受信モジュール28を介してPCIブリッジモジュール26-2にパケット転送する際に、アドレス40-2に変換される。

【0036】アドレス40-2は、ノードID44とMSUアドレスオフセット46で構成される。この場合、ノードID44は同様にメモリモジュール32-2に対応したPCIブリッジモジュール26-2のノードIDである。またMSUアドレスオフセット46は、同じくメモリモジュール32-2に設けている主記憶ユニット34-2のメモリ空間のアドレスを示すオフセットである。

【0037】更にPCIブリッジモジュール26-2からメモリモジュール32-2に出力するアドレス40-3は、PCIバス30-2に接続したメモリモジュール32-2を指定するアドレス42と、メモリモジュール32-2におけるMSUアドレスオフセット46で構成される。

【0038】再び図3を参照するに、このような処理モジュール20-1からパケットネットワークを経由したライトコマンドの転送によるメモリモジュール32-2のライトアクセスに対し、本発明にあっては更に、処理モジュール20-1に検証コマンド生成制御部36を設け、またこれに対応したPCIブリッジモジュール24

-1に検証コマンド応答部38を設けている。

【0039】処理モジュール20-1に設けた検証コマンド生成制御部36は、処理モジュール20-1がPCIブリッジモジュール24-1に対しライトコマンドを発行してPCIバス12-1によるコマンド転送を正常終了した後に、パケットネットワークによるPCIブリッジモジュール26-2へのライトコマンドのパケット転送状態を検証する検証コマンドを発行する。

【0040】PCIブリッジモジュール24-1に設けた検証コマンド応答部38は、処理モジュール20-1に設けた検証コマンド生成制御部36からの検証コマンドを受信した際に、ライトコマンドのパケット転送に対応して転送先のPCIブリッジモジュール26-2から受信しているリブライパケットをリードデータとして応答することで、パケット転送によるデータ書込みが正常に行われたか否かを処理モジュール20-1に認識させる。

【0041】処理モジュール20-1に設けた検証コマンド生成制御部36で発行する検証コマンドはリードコマンドであり、このリードコマンドのアドレスとして書込先のメモリモジュール32-2の実アドレス空間を越える仮想アドレス空間を指定したリードコマンドを検証コマンドとして発行する。

【0042】図5は処理モジュール20-1から見たメモリモジュール32-2のメモリ空間60の説明図である。このメモリ空間60は、処理モジュール20-1のアクセス可能領域66と処理モジュール20-2のアクセス可能領域68を実アドレス空間62として持っており、アクセス可能領域66はページ0～n-1で管理され、またアクセス可能領域68はページn～2n-1で管理されている。

【0043】このような処理モジュール20-1から見たメモリモジュール32-2の実アドレス空間62に対し、検証コマンド生成制御部36で検証コマンドとして発行するリードコマンドのアドレスは、実アドレス空間62を越える所定のオフセットを加算した仮想アドレス空間64のアドレスを指定する。

【0044】ここで実アドレス空間62における処理モジュール20-1のアクセス可能領域66における先頭アドレスを「aaaa」とし、また処理モジュール20-2のアクセス可能領域68の先頭アドレスを「bbbb」とすると、実アドレス空間62のアドレスを使用したライトコマンドに続いて発行される検証コマンドのアドレスは所定のオフセット「cccc」を加算した仮想アドレス空間64のアドレスとなる。

【0045】したがって検証コマンドによりアクセスされる仮想アドレス空間64は、処理モジュール20-1についてはアドレス「aaaa+cccc」を先頭アドレスとしたページ0～n-1の検証アクセス領域70となる。また処理モジュール20-1については先頭アド

レス「bbbb+cccc」からのページ $n \sim 2n-1$ を割り当てた検証アクセス領域72となる。

【0046】この検証アクセス領域70、72となる仮想アドレス空間64は、通常の実アドレス空間62に対するリードコマンドでは使用されないユニークなアドレスである。このためPCIブリッジモジュール24-1の検証コマンド応答部38は処理モジュール20-1からのリードアクセスのアドレスを判別し、アドレスが仮想アドレス空間64に存在することを認識した場合は、既に発行しているライトコマンドに対する検証コマンドであることを認識し、このとき転送先のPCIブリッジモジュール26-2から受信しているリブライバケットをリードデータとして処理モジュール20-1に回答する。

【0047】このため、PCIブリッジモジュール24-1からPCIブリッジモジュール26-2にライトコマンドをバケット転送した後の正常終了または異常終了を示すリブライバケットのデータを、PCIバス12-1のハードウェア構成や機能を変更することなく、リードコマンドを使用したエミュレーションにより処理モジュール20-1で確実に認識することができ、バケットネットワークによるライトコマンドのバケット転送で生じた異常をPCIバス12-1側の処理モジュール20-1で認識して、適切なエラーリカバリをとることができる。

【0048】図6は図3の処理モジュール20-1からバケットネットワークを経由してメモリモジュール32-2にライトコマンドを発行した書込処理のタイムチャートである。

【0049】まず処理モジュール20-1がホスト10-1からのライト命令に基づいて、ステップS1でメモリモジュール32-2に対するライトコマンドを発行する。この場合のライトコマンドのアドレスは、図5の処理モジュール20-1から見たメモリ空間60のアドレス「aaaa」を指定している。PCIバス12-1からPCIブリッジモジュール24-1に対するライトコマンドのPCIバス12-1による転送が正常終了すると、処理モジュール20-1とPCIブリッジモジュール24-1のバス接続は切り離される。

【0050】ライトコマンドを受信したPCIブリッジモジュール24-1は、アドレスに含まれる送信先のノードIDに基づき、バケット送受信モジュール28を介してPCIブリッジモジュール26-2に対し、ステップS2でライトバケットを送信する。

【0051】PCIブリッジモジュール26-2はPCIブリッジモジュール24-1からのライトバケットをステップS3で受信し、受信バケットから得られたライトコマンドをメモリモジュール32-2にPCIバス12-2により転送する。このためメモリモジュール32-2は、ステップS4で受信したライトデータを主記憶

ユニット34-2に書き込む。

【0052】一方、PCIブリッジモジュール26-2はステップS3でライトバケットを受信し、その受信状態即ち正常終了から異常終了かに応じ、この場合には正常終了であることから、ステップS5でリブライバケットを送信元のPCIブリッジモジュール24-1に送信する。

【0053】PCIブリッジモジュール24-1は、ステップS2でライトバケットを送信した後、ステップS7で送信先のPCIブリッジモジュール26-2からのリブライバケットを受信して保存する。

【0054】一方、処理モジュール20-1にあっては、ステップS1でライトコマンドを発行した後、ステップS6でアドレス「aaaa+cccc」を指定したリードコマンドを検証コマンドとしてPCIブリッジモジュール24-1にバス転送する。

【0055】このステップS6のリードコマンドとしての検証コマンドを受信したPCIブリッジモジュール24-1は、そのアドレス「aaaa+cccc」により、図5に示した検証アクセス領域70のリードコマンドであることを認識し、ステップS7で受信しているPCIブリッジモジュール26-2からのリブライバケットをリードデータとして処理モジュール20-1に送信する。

【0056】処理モジュール20-1は、ステップS6の検証コマンドに対する応答として、PCIブリッジモジュール24-1からのリブライデータをステップS9で受信し、リブライデータを解釈して正常終了か異常終了かを認識することができる。

【0057】ステップS1でライトコマンドを発行した後のステップS6の検証コマンドとしてのリードコマンドの発行は、PCIブリッジモジュール24-1、26-2間におけるライトバケットの転送とは非同期である。

【0058】このためライトバケットの転送中にステップS6のリードコマンドが発行されることもあり、この場合にはライトバケットの転送中であることから、PCIブリッジモジュール24-1はリトライコマンドを応答し、これにより処理モジュール20-1からの検証コマンドとしてのリードコマンドの発行を繰り返させ、転送終了後に、最初に受け取った検証コマンドとしてのリードコマンドに対しリブライデータを応答することになる。

【0059】図7は、図3の処理モジュール20-1に設けた検証コマンド生成制御部36の具体的な回路ブロック図である。処理モジュール20-1に設けた検証コマンド生成制御部36は、インタフェース制御部48、検証アクセスコマンド生成部50、検証アクセス結果バッファライト制御部52、検証アクセス結果格納バッファ54、検証アクセス結果チェック部56及びエラー通

11

知レジスタ58で構成される。

【0060】検証アクセスコマンド生成部50はライトコマンド発行後に動作し、ライトコマンドのアドレスに所定のオフセット「cccc」を加算したアドレスのリードコマンドを検証コマンドとして、インタフェース制御部48によりPCIバス12-1を介してPCIブリッジモジュール24-1に発行する。

【0061】検証アクセスコマンド生成部50から検証コマンドが発行されると、検証アクセス結果バッファライト制御部52が動作し、インタフェース制御部48で受信したリブライバケットを検証アクセス結果格納バッファ54に格納する。検証アクセス結果格納バッファ54に格納されたリブライバケットは検証アクセス結果チェック部56で解析され、正常終了か異常終了かが判別される。

【0062】異常終了が判別された場合にはエラー通知レジスタ58にエラー登録が行われ、処理モジュール20-1の検証機能によってエラー通知レジスタ58のエラー登録が認識されると、ライトコマンドの再発行などのリカバリが行われることになる。

【0063】図8は、図3のPCIブリッジモジュール24-1のブロック図である。このPCIブリッジモジュール24-1は、PCIインタフェース制御部24、アドレス/ページ変換部76、バッファライト制御部78、バッファステータス保持部80、バッファ選択制御部82、エラー検出バッファ保持機能部84、バッファ86-1、86-2、86-3、86-4、インタフェース制御部88、リブライバケット解析部90、ページ/バッファ番号参照部92、リブライバケット参照部94を備える。

【0064】このような回路部はPCIバスとバケットネットワークの接続変換に使用する基本的な回路部を構成する。これに加え本発明にあっては、更に、検証コマンド応答部38を設けている。検証コマンド応答部38には検証アクセスコマンド検出部96、アドレス/ページ変換部98及びバケット転送中登録部100が設けられる。

【0065】図8のPCIブリッジモジュール24-1の詳細は、図9及び図10に分割して示すようになる。図9及び図10にあっては、各ブロックとの間の制御線及びデータ線の様子が詳細に示されている。

【0066】この図8乃至図10に示した図3のPCIブリッジモジュール24-1のライトコマンドに基づくメモリ書込処理を、図11、図12に分割して示すタイムチャートを参照して説明すると次のようになる。

【0067】図11において、処理モジュール20-1は、ステップS1でメモリモジュール32-2に対するライトコマンドを発行する。このライトコマンドはメモリモジュール32-2のメモリ空間60におけるページ0へのPCIライトであり、図4に示したアドレス40

12

-1を含むライトコマンドがPCIバス12-1によりPCIブリッジモジュール24-1に転送される。

【0068】PCIブリッジモジュール24-1のアドレス/ページ変換部76は、ステップS2で受信アドレスをページに変換し、例えばページ0と判別する。続いてバッファライト制御部78が、判別されたページ0へのバッファ割当てをバッファ選択制御部82に要求する。

【0069】このためバッファ選択制御部82は、ステップS4でバッファステータス保持部80-1、80-2、80-3、80-4を参照し、使用可能なバッファをバッファライト制御部78に通知する。バッファライト制御部78はステップS6で、バッファ選択制御部82から通知されたバッファに処理モジュール20-1から受信したデータをライトした後、インタフェース制御部88に対しPCIブリッジモジュール26-2への送信を要求する。

【0070】これによりインタフェース制御部88は、処理モジュール20-1からのデータをライトしたバッファ86-1〜86-4のいずれかを選択し、そのデータを読み出してバケット変換した後、バケット送受信モジュール28を介してPCIブリッジモジュール26-2にバケットを送信する。

【0071】ステップS6でバッファライト制御部78がPCIブリッジモジュール26-2へのバケット送信を要求すると、ステップS7で、通知されたバケット番号と対応するページ番号を、ページ/バッファ番号参照部92に内蔵したテーブルに登録する。

【0072】このときインタフェース制御部88からはステップS8のように、PCIブリッジモジュール26-2へのデータ送信後、そのバケット番号をページ/バッファ番号参照部92に通知する。したがってページ/バッファ番号参照部92にあっては、ステップS9で、通知されたバケット番号と対応するバケット番号を内蔵テーブルに登録する。

【0073】次に図12のように、インタフェース制御部88において送信先のPCIブリッジモジュール26-2からのリブライバケットの受信が行われると、ページ/バッファ番号参照部92及びリブライバケット参照部94に対し受信通知が行われる。

【0074】ページ/バッファ番号参照部92は、ステップS11で、通知されたバケット番号から該当ページを参照し、リブライバケット参照部94に通知する。リブライバケット参照部94はインタフェース制御部88から通知されたバケット番号に対応するページ/バッファ番号参照部92から通知されたページ番号を内蔵テーブルにステップS12で登録する。

【0075】バケット転送中登録部100は、図11のステップS5でページ0の転送処理中の登録を行っているが、ステップS13でリブライバケット参照部94よ

り登録終了通知を受けると、ページ0を転送処理中としていた登録を解除する。

【0076】一方、処理モジュール20-1にあっては、図11のステップS1でライトコマンドを発行した後、図12のステップS15で検証コマンドとしてのリードコマンドを発行する。このリードコマンドは、ライトコマンドのページ0のアドレスに所定のオフセット「cccc」を加算した検証アクセス領域のアドレスを指定したリードコマンドである。

【0077】ステップS15で処理モジュール20-1が発行した検証コマンドとしてのリードコマンドは、ステップS16のようにPCIブリッジモジュール24-1の検証アクセスコマンド検出部96で受信され、アドレスが検証アクセス領域にあることが判別され、検証コマンドであることが認識される。

【0078】そこで検証アクセスコマンドの検出結果がアドレス/ページ変換部98に出力され、ステップS17でページ0への検証アクセスを要求する。この検証アクセスの要求は、ステップS14でバケット転送中登録部100を参照してページ0の転送終了を確認した後、ステップS18でリブライバケット参照部94に保持しているページ0に対応するリブライバケットを読み出して、処理モジュール20-1にリードデータとして送信する。

【0079】尚、アドレス/ページ変換部98から検証アクセスの要求を受けた際に、ライトコマンドの転送中の登録状態であった場合には、バケット転送中登録部100はPCIインタフェース制御部74に対しリトライ要求コマンド(Retry REQ.)を応答し、処理モジュール20-1は転送終了まで検証コマンドとしてのリードコマンドの発行を繰り返す。

【0080】このような処理モジュールからバケットネットワークを経由してメモリモジュールにデータを書き込むライトアクセスは、図3の処理モジュール20-1、PCIブリッジモジュール24-1、バケット送受信モジュール28、PCIブリッジモジュール26-2及びメモリモジュール32-2となる経路のアクセスを例にとるものであったが、処理モジュール20-1からPCIブリッジモジュール26-1に対応したメモリモジュール32-2にアクセスする場合も、アドレスのメモリ空間が異なる以外は同様にして行われる。

【0081】更に、ホスト10-2に対応して設けている処理モジュール20-2にも本発明による検証コマンド生成制御部36が設けられ、処理モジュール20-2に対応したPCIブリッジモジュール24-2にも本発明による検証コマンド応答部38が設けられている。

【0082】この処理モジュール20-2もPCIブリッジモジュール24-2からバケットネットワークを経由してPCIブリッジモジュール26-1のメモリモジュール32-1またはPCIブリッジモジュール26-

2のメモリモジュール32-2にライトアクセスを同様にして行うことができる。

【0083】更にバケット送受信モジュール28は、ネットワーク18により図2に示した他の筐体16-2～16-nにも接続されており、他の筐体に設けているメモリモジュールに対しても、そのメモリ空間のアドレスを指定することで同様にしてライトコマンドのバケット転送による書込みができる。

【0084】この場合にも他の筐体のメモリモジュールに対応して設けているPCIブリッジモジュールからのリブライバケットをPCIブリッジモジュール24-1または24-2で受信し、処理モジュール20-1または20-2の検証コマンド生成制御部36からの検証コマンドに対しリブライデータをリードデータとして応答することで、他の筐体に対しバケット転送で行ったライトアクセスが正常終了か異常終了かを確実に認識することができる。

【0085】尚、本発明は上記の実施形態に限定されず、その目的と利点を損なわない適宜の変形を含む。また、本発明は上記の実施形態に示した数値による限定は受けない。

【0086】

【発明の効果】以上説明してきたように本発明によれば、PCIバス等のインタフェースバスの処理モジュールからバケットネットワークのブリッジモジュールへのライトコマンドの転送による正常終了後のバケット転送の結果を示すリブライバケットにつき、ライトコマンドのアドレス空間を越えるユニークなアドレス指定によるリードコマンドを検証コマンドとして発行することで、リードコマンドの応答データとしてリブライバケットをPCIバスの処理モジュールで取得することができ、バケットネットワークにおける転送状態をPCIバス側の処理モジュールで確認することができ、バケット転送が異常終了した場合にはライトコマンドの再発行などの適切な対応策をとることができ、バケットネットワークを経由した主記憶アクセスの信頼性を向上することができる。

【0087】またバケットネットワークで受信されたリブライバケットのPCIバス側の処理モジュールでの認識について、PCIバスで使用している通常のリードコマンドを検証コマンドとして使用しているため、PCIバス側の構成や制御手順を何ら変更することなく、確実にバケットネットワーク側で受信したリブライバケットをPCIバス側の処理モジュールで認識することができる。

【図面の簡単な説明】

【図1】本発明の原理説明図

【図2】本発明が適用される計算機システムのブロック図

【図3】図2の筐体に収納された本発明によるメモリ制

15

御装置の実施形態を示したブロック図

【図4】図3の処理モジュール、送信側のブリッジモジュールおよび受信側のブリッジモジュールの各々におけるアドレス変換の説明図

【図5】図2の処理モジュールから検証コマンドとして発行するリードコマンドによるメモリ空間のアドレスの説明図

【図6】図2の実施例におけるライトアクセスのタイムチャート

【図7】図3の処理モジュールに設けた検証コマンド発行制御部の実施形態のブロック図

【図8】本発明による検証コマンド応答部を備えた図3のPCIブリッジモジュールのブロック図

【図9】図8のPCIブリッジモジュールの詳細ブロック図

【図10】図9に続くPCIブリッジモジュールの詳細ブロック図

【図11】図9、図10のPCIブリッジモジュールにおけるライトアクセス動作のタイムチャート

【図12】図11に続くPCIブリッジモジュールにおけるライトアクセス動作のタイムチャート

【図13】PCIバスをバケットネットワークを介して接続した際のメモリアクセスで生ずる問題の説明図

【符号の説明】

10, 10-1, 10-2: ホスト

12-1, 12-2, 30-1, 30-2: PCIバス
(インタフェースバス)

14-1, 14-2: デバイス

16-1~16-n: 筐体

18: バケットネットワーク

20, 20-1, 20-2: 処理モジュール

16

* 22-1, 22-2: デバイスモジュール

24, 24-1, 24-2: PCIブリッジモジュール
(送信側)

26, 26-1, 26-2: PCIブリッジモジュール
(受信側)

28: バケット送受信モジュール

32-1, 32-2: メモリモジュール

34-1, 34-2: 主記憶ユニット (MSU)

36: 検証コマンド発生制御部

38: 検証コマンド応答部

50: 検証アクセスコマンド生成部

52: 検証アクセス結果バッファライト制御部

54: 検証アクセス結果格納バッファ

56: 検証アクセス結果チェック部

58: エラー通知レジスタ

60: メモリ空間

62: 実アドレス空間

64: 仮想アドレス空間

74: PCIインタフェース制御部

76: アドレス/ページ変換部

78: バッファライト制御部

80: バッファステータス保持部

82: バッファ選択部

84: エラー検出バッファ機能部

86-1~86-4: バッファ

90: リブライバケット解析部

92: ページ/バッファ番号参照部

94: リブライバケット参照部

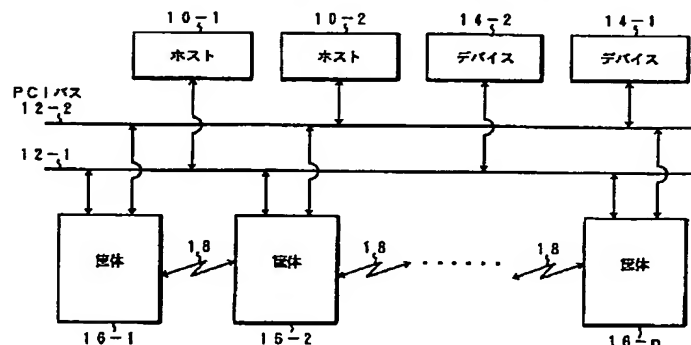
96: 検証アクセスコマンド検出部

98: アドレス/ページ変換部

* 100: バケット転送中登録部

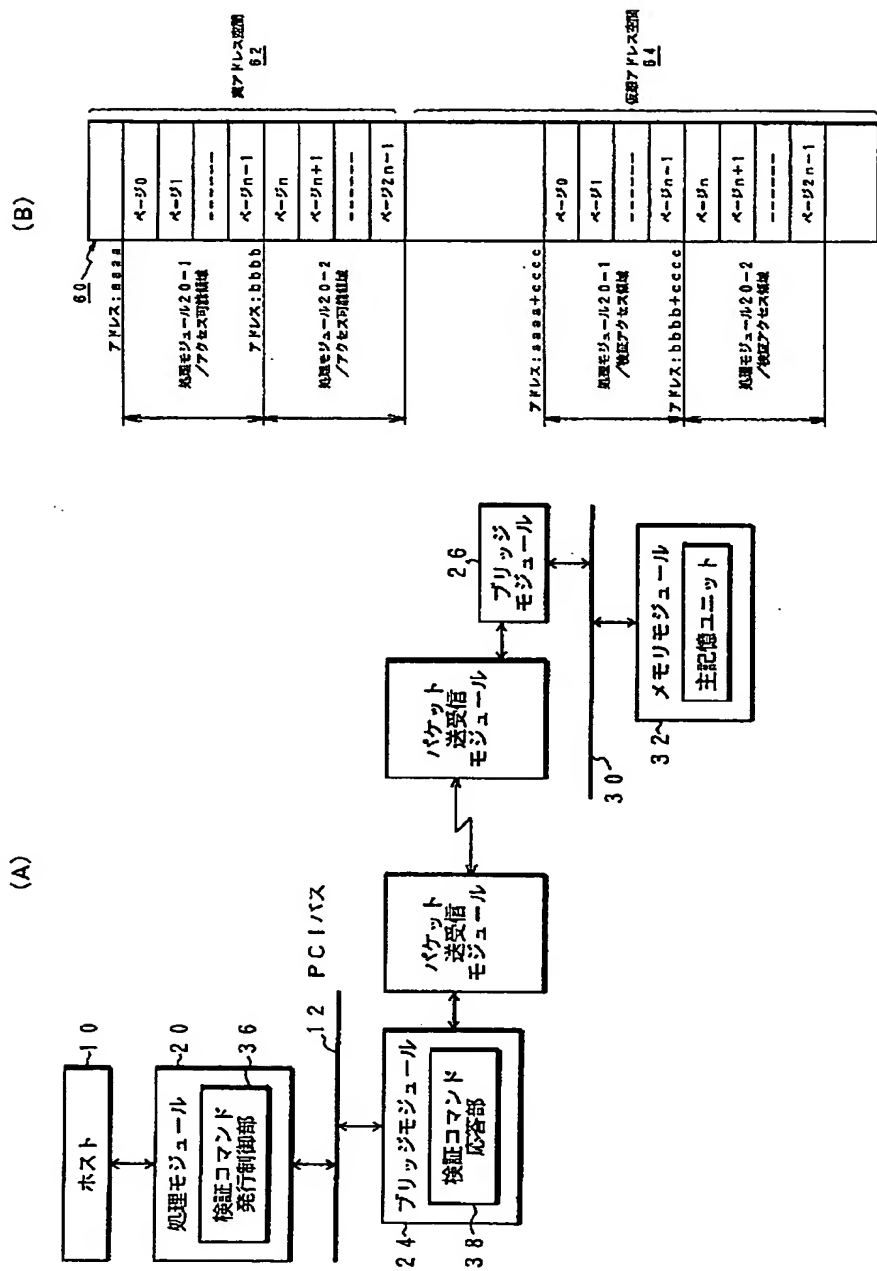
【図2】

本発明が適用される計算機システムのブロック図

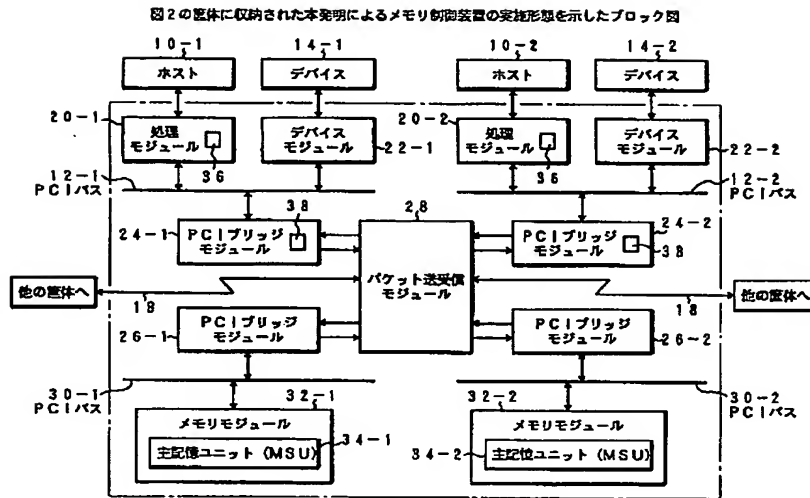


【図1】

本発明の原理説明図



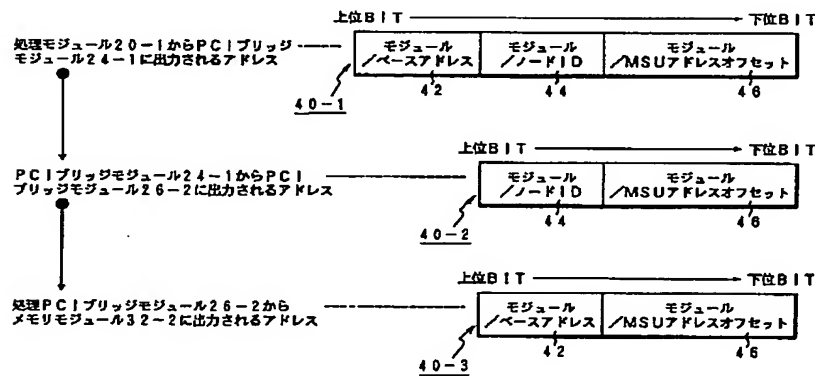
【図3】



【図4】

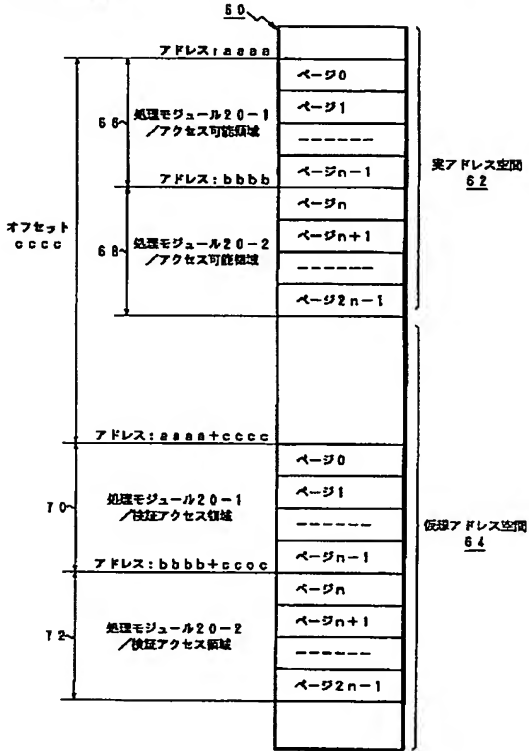
図3の処理モジュール、送信側のブリッジモジュールおよび受信側のブリッジモジュールの各々におけるアドレス変換の説明図

モジュールA1からモジュールEへのアクセスの場合



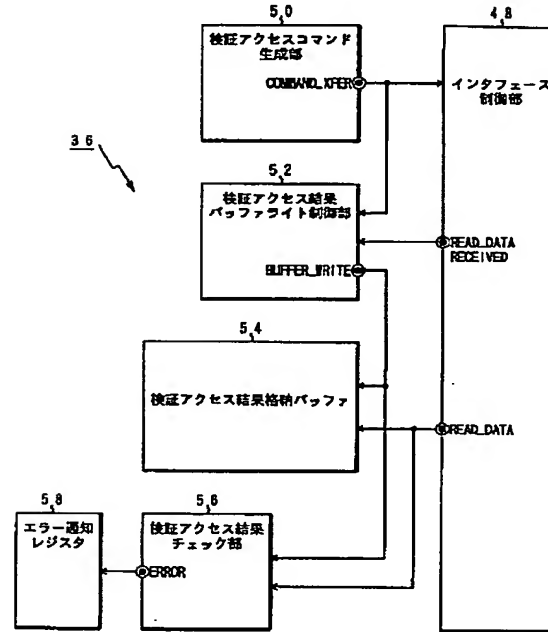
【図5】

図2の処理モジュールから検証コマンドとして発行するリードコマンドによるメモリ空間のアドレスの配列図



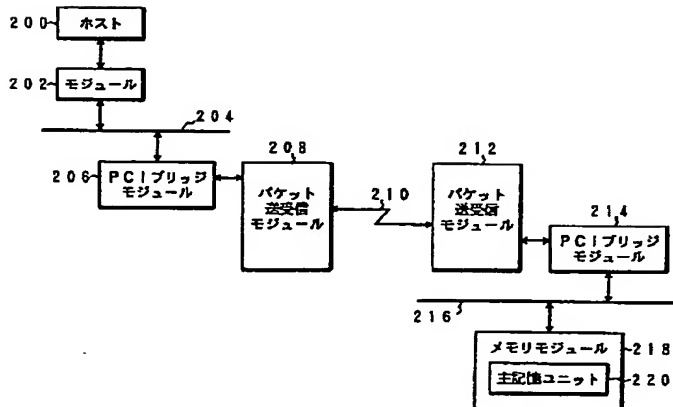
【図7】

図3の処理モジュールに設けた検証コマンド発行制御部の実施形態のブロック図



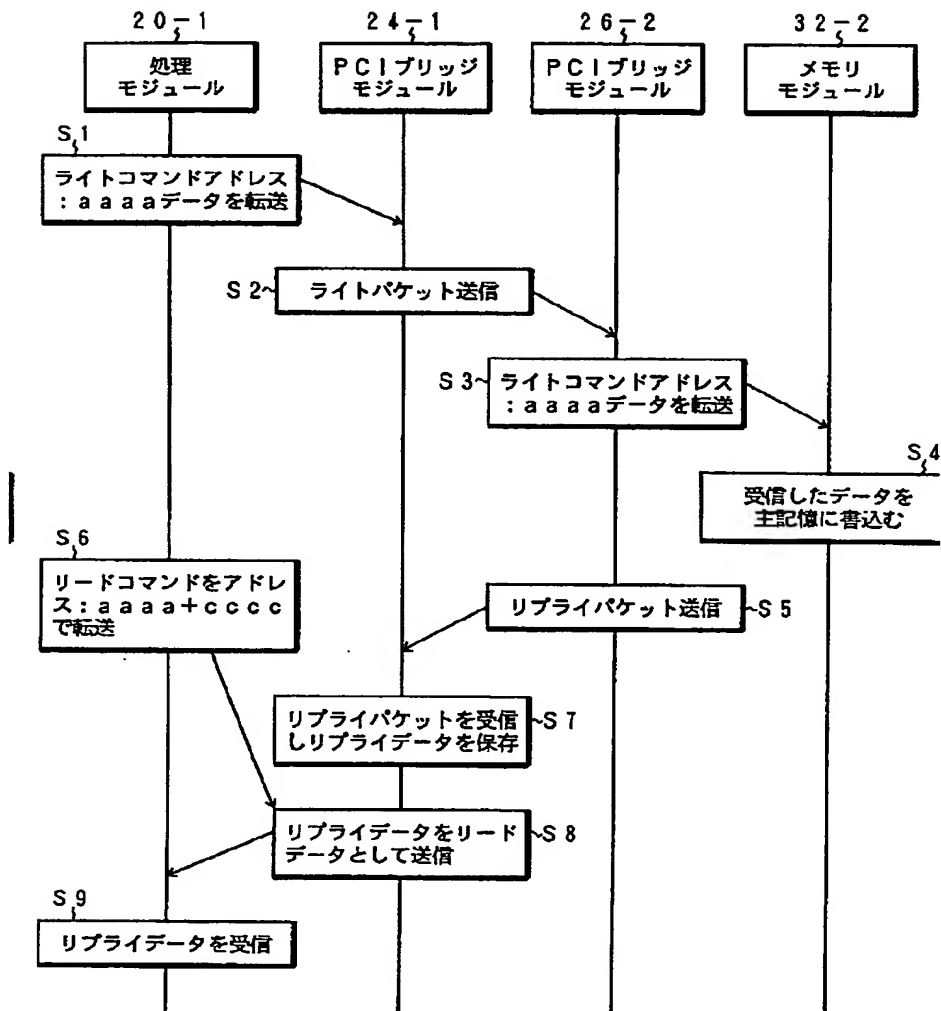
【図13】

PCIバスをパケットネットワークを介して接続した際のメモリアクセスで生ずる問題の説明図



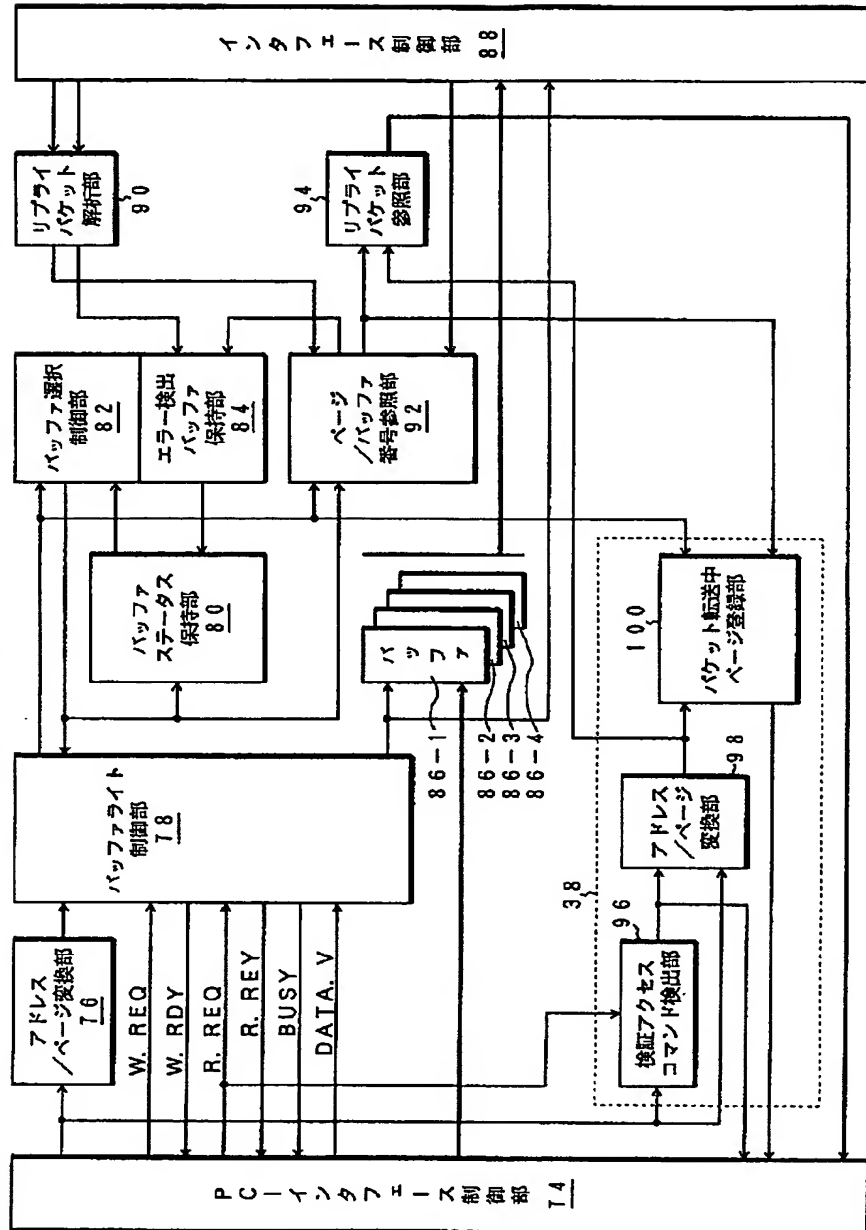
【図6】

図2の実施例におけるライトアクセスのタイムチャート



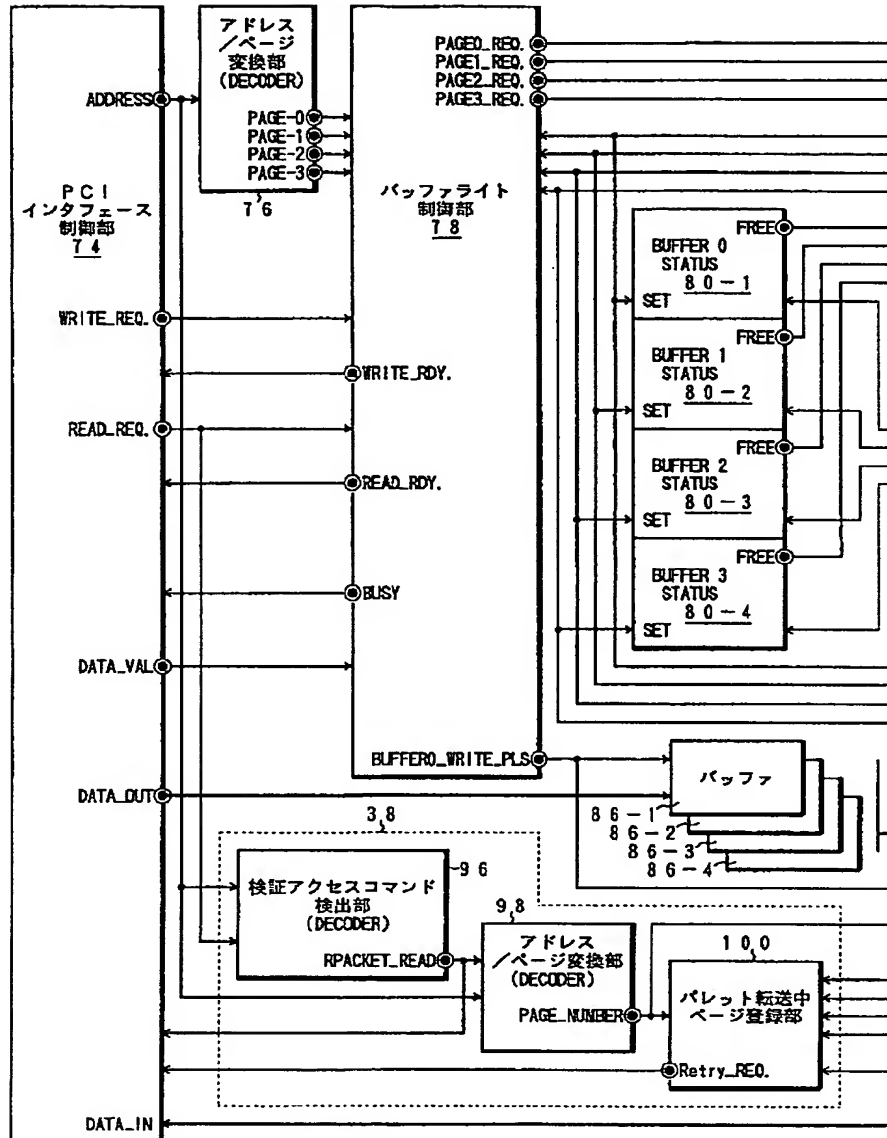
【図8】

本発明による検証コマンド応答部を備えた図3のPCブリッジモジュールのブロック図



【図9】

図8のPCIブリッジモジュールの詳細ブロック図



【図10】

図9に換くPCIブリッジモジュールの詳細ブロック図

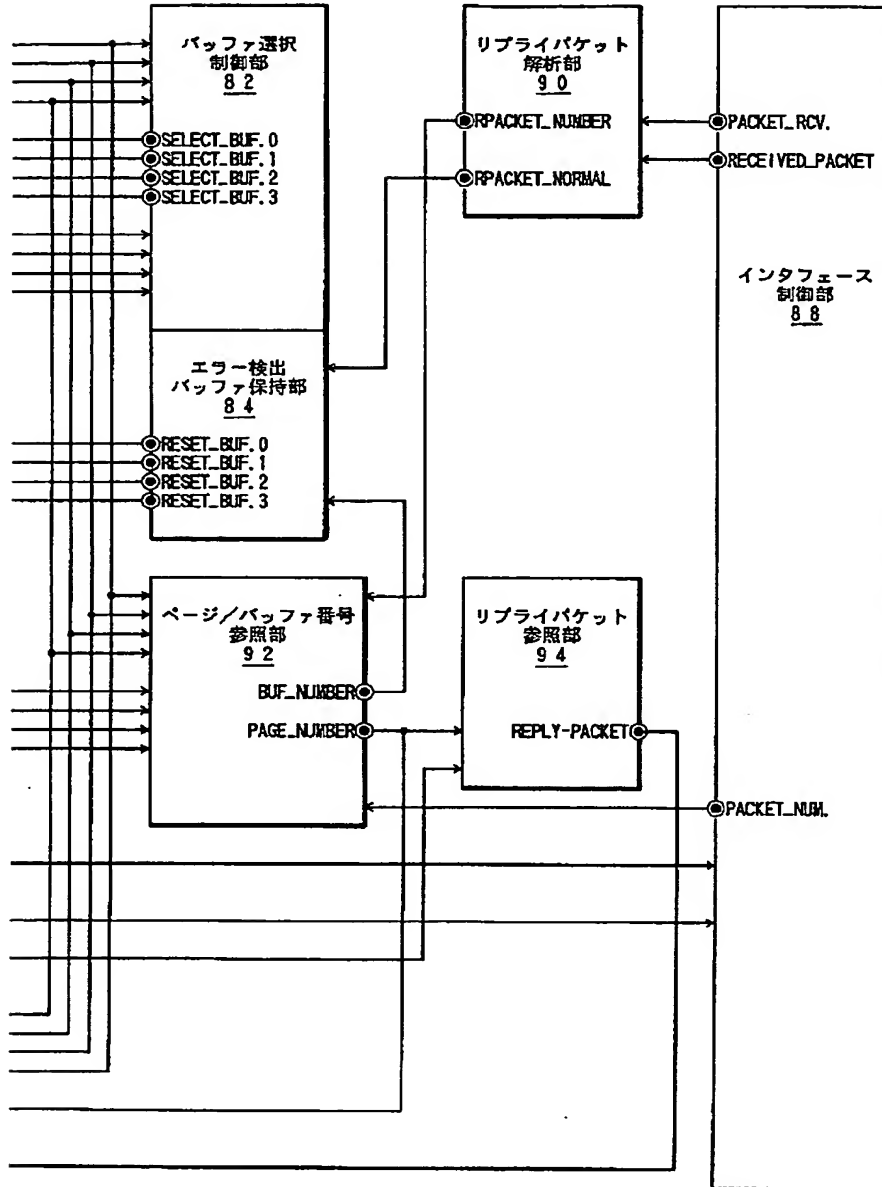
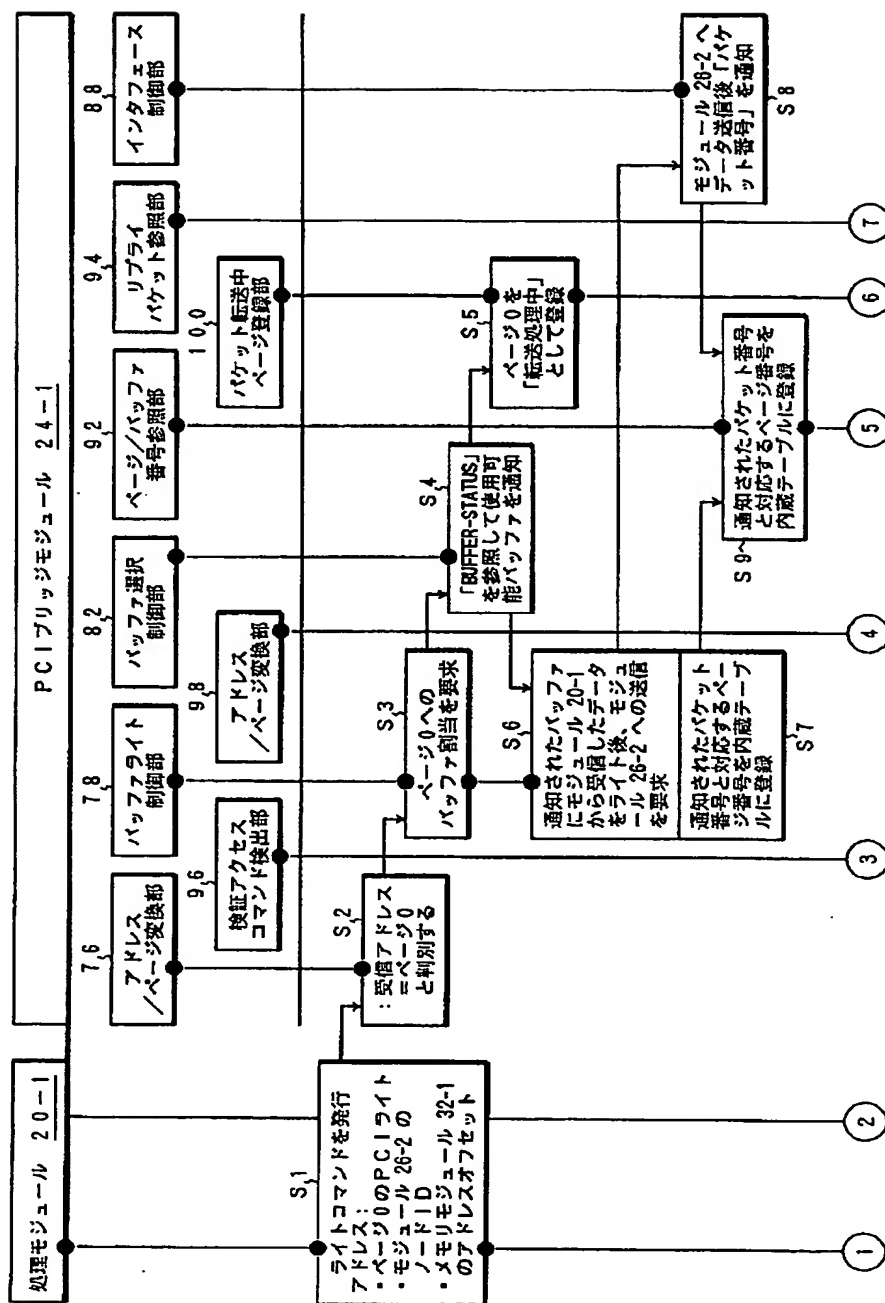


図9、図10のPC・ブリッジモジュールにおけるライトアクセス動作のタイムチャート



(72)発明者 富張 清隆

神奈川県横浜市港北区新横浜 2 丁目 4 番 19

号 株式会社富士通プログラム技研内

F ターム (参考)

5B018 GA07 HA31 KA02 MA01 QA16

5B045 BB12 BB29 BB48 DD01

5B060 CA00

5B083 AA08 BB02 CD00 CE00 EE03

EF11 GG04

5K034 DD03 EE11 HH42